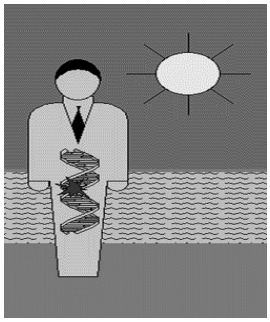
Whole Genome Association Study and Risk Assessment

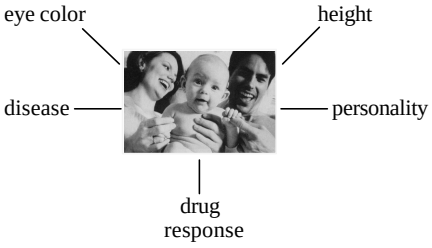
David R. Cox M.D., Ph.D.
Chief Scientific Officer
Perlegen Sciences Inc.

david_cox@perlegen.com

Genes, through the proteins they encode, interact with challenges from the environment



Variations in DNA sequence are responsible for all inherited differences between people



Single Nucleotide Polymorphisms (SNPs)

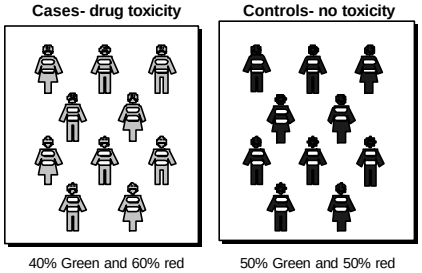
- DNA sequence comparison of any two copies of the human genome reveals only 0.1% sequence variability
- Each variable base (SNP) results from a single error in DNA replication that occurred once in the history of mankind
- Each SNP is characterized by only two bases
- Common variants are more experimentally tractable than rare variants

SNPs result in functional differences by altering the quality and/or the quantity of cellular proteins

Why has it been so difficult to identify genes determining disease or drug response?

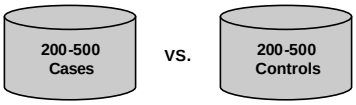
- **Most common human traits are not caused by a single variation, but probably by <u>20 or more</u> genetic changes across the genome**
- **Any single genetic change may be responsible for only a small contribution to the trait**

Genetic Association Analysis



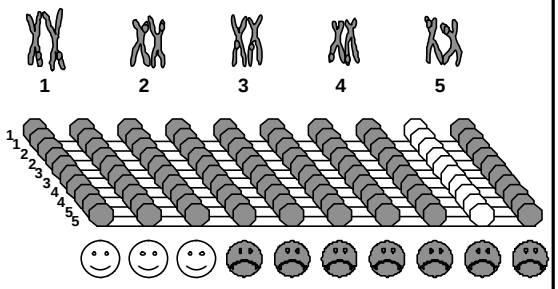
If a segment of the genome is "associated" with toxicity, cases will have a different SNP allele frequency than controls.

Genetic Association Studies



- **Obtaining sufficient statistical power to find**
- **genes with small effects requires analyzing the DNA of large numbers of people**:

A single common disease or drug response results from different combinations of variants in a limited number of genes



Realistic Genetic Opportunities

Pharmacogenetics: Identify the set of genetic markers across the human genome that determine variation in drug efficacy and safety

Disease Genetics: Improve the efficiency of clinical trails by defining the biologic basis of disease

What is Needed for a Comprehensive Study of the Relationship Between Common SNPs and Drug Response or Disease Susceptibility?

- Human SNP discovery
- High-throughput low-cost genotyping assays
- Well defined measures of drug response and disease
- Variation in drug response and disease between individuals

Perlegen Sciences Achieved A Major Breakthrough in Human DNA Variation Technology in 2001

Blocks of Limited Haplotype Diversity Revealed by High-Resolution Scanning of Human Chromosome 21

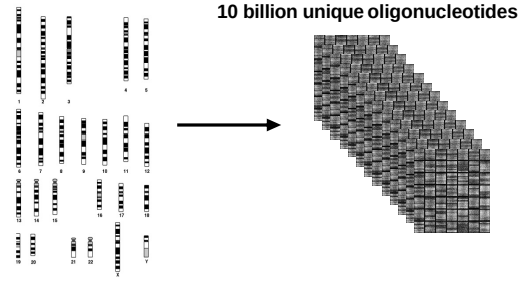
Nila Patil, Anthony J. Berno, David A. Hinds, Wade A. Barrett, Jigna M. Doshi, Coleen R. Hacker, Curtis R. Kautzer, Danny H. Lee, Claire Marjoribanks, David P. McDonough, Bich T. N. Nguyen, Michael C. Norris, John B. Sheehan, Naiping Shen, David Stern, Renee P. Stokowski, Daryl J. Thomas, Mark O. Trulson, Kanan R. Vyas, Kelly A. Frazer, Stephen P. A. Fodor, David R. Cox*

SCIENCE VOL 294 23 NOVEMBER 2001

Perlegen Sciences Discovered 1.6 Million Human SNPs Using High-Density Oligonucleotide Arrays

Human genome
3 billion base pairs

A collection of 223 high-density arrays containing more than 10 billion unique oligonucleotides



Perlegen's Genotyping Platform

- **Perlegen has designed special high-density arrays which measure 1.6 million unique SNPs …**

Perlegen's Individual Genotyping Process

One hybridization yields genotypes of 85,000 SNPs in a single individual

Over the past 12 month Perlegen has produced more than 2 billion genotypes

What are the relationships between common SNPs in different human populations?

Perlegen Sciences genotyped 1,586,358 SNPs in 23 African American, 24 European American, and 24 Han Chinese unrelated individuals over a 6 week period

All 112 million genotypes are deposited in build 123 of dbSNP

Whole-Genome Patterns of Common DNA Variation in Three Human Populations

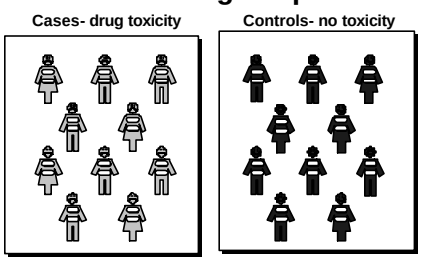
David A. Hinds,[1] Laura L. Stuve,[1] Geoffrey B. Nilsen,[1] Eran Halperin,[2] Eleazar Eskin,[3] Dennis G. Ballinger,[1] Kelly A. Frazer,[1] David R. Cox[1]*

***Science*, Vol 307, 1072-1079 , 18 February 2005**

Whole-Genome Association Analysis

Genotypes of sets of 300000- 500000 "tag SNPs" provide information regarding a large fraction of all 6 million common SNPs present in human populations.

Genetic Association Analysis of An Adverse Drug Response

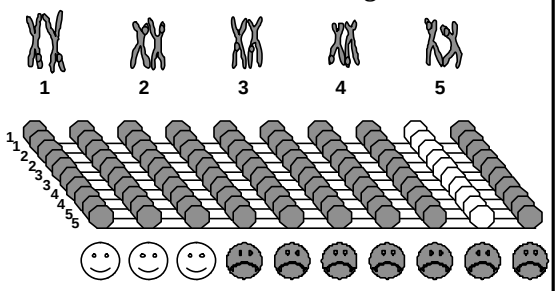


If a segment of the genome is "associated" with toxicity, cases will have a different SNP allele frequency than controls.

Individual SNP Relative Risk

Call rate	Delta P	P_CHISQ	heterozygote RR	homozygote RR
0.93	-11.6%	8.0E-06	2.0	4.8
0.99	14.7%	1.8E-06	1.4	3.1
0.99	12.2%	4.8E-06	1.8	2.4
0.99	12.4%	6.4E-06	1.6	2.3
1.00	13.9%	7.4E-06	1.1	2.0
0.98	10.3%	8.7E-06	3.4	6.5
0.93	-11.6%	8.0E-06	2.0	4.8
0.98	10.0%	7.7E-06	5.8	12.4
1.00	13.9%	7.4E-06	1.1	2.0
0.99	10.9%	6.4E-06	2.7	5.7
0.99	12.4%	6.4E-06	1.6	2.3
0.91	10.4%	5.6E-06	5.3	11.9
0.99	12.2%	4.8E-06	1.8	2.4
0.99	10.5%	3.5E-06	6.4	13.7
0.99	14.7%	1.8E-06	1.4	3.1

A single common disease or drug response results from different combinations of variants in a limited number of genes



Using A 34 SNP "Barcode" to Identify Patients At Significantly Increased Risk Of An Adverse Event

# alleles	controls	cases	barcode RR
>45	1.1%	23.0%	21.6
>44	1.6%	31.1%	19.5
>43	2.7%	36.6%	13.8
>42	4.3%	44.3%	10.4
>41	7.4%	49.2%	6.6
>40	11.7%	55.7%	4.8

How will this new genetic information be applied in general medical practice?

The Missing Piece

A national clinical network assessing treatment outcomes for a wide range of disorders

Genetic association analysis of data collected by such a network would provide an important scientific body of knowledge that could be used to improve treatment efficacy and to reduce adverse treatment events in individual patients